A Keystone for ncRNA

Genome Biology

MEETING REPORT

A Keystone for ncRNA

Ezgi Hacisuleyman[1,2,3], Moran N Cabili[2,3,4] and John L Rinn[2,3]*

Abstract

A report on the Keystone symposium 'Non-coding RNAs' held at Snowbird, Utah, USA, 31 March to 5 April 2012.

Keywords Non-coding RNA, epigenetics, chromatin

Once upon a time RNA fit cleanly into the central dogma as a messenger between DNA and protein. Over the past 50 years, RNA molecules have continually emerged as dynamic and versatile regulators of the genome. Our modern understanding of non-coding RNAs (ncRNAs) may look like an intertwined mess of molecules, but collectively they exhibit architecture and coordination, leading to elegantly choreographed regulation of DNA and protein by RNA. The emerging role of RNA as an orchestrator resonated throughout the inaugural Keystone symposium on ncRNAs, with many examples of long ncRNAs (lncRNAs) having critical roles across numerous biological pathways. After the meeting, it was clear that more surprises would emerge from the 'dark matter' of the genome.

The keynote address by Nick Proudfoot (University of Oxford) set the stage by describing how the most classically studied regions of the genome, such as the β-globin locus, are now emerging as being exquisitely regulated by RNA. Specifically, Proudfoot summarized years of work showing how the act of transcription through non-coding regions, and importantly where transcriptional termination occurs, regulates the epigenetic dynamics of the locus. Intriguingly, convergent transcription by RNA polymerase II (RNA pol II) may serve as a substrate to recruit Dicer and other factors of the RNA interference (RNAi) machinery. Similarly, Robert Martienssen (Cold Spring Harbor Laboratory) presented an interplay between RNA/DNA polymerase activity and RNAi in establishing heterochromatic domains. The dependence on co-transcriptional RNAi allows the release of RNA polymerase and prevents collision with the centromeric DNA replication machinery. Together these studies demonstrate the need for not only identifying lncRNAs involved in epigenetic establishment but also for understanding many simultaneous intertwined layers of regulation.

The human noncoding transcriptome reveals a map of 'noncodarnia'

Thomas Gingeras (Cold Spring Harbor Laboratory) provided an overview of the complexity of the human transcriptome resulting from the efforts of the ENCODE consortium. The transcriptomic map has gained an unprecedented resolution, revealing that 76% of our genome is transcribed. With an average of approximately eight transcripts per genic region, the wealth of ENCODE has redefined the 'one gene - one function' hypothesis into 'many transcripts - one function', or possibly many. Using complementary datasets and approaches, Piero Carninci and the Riken OMICs Center have provided new insights into lncRNA promoter regulation. By fine mapping of the 5' 7-methyl guanosine caps on RNA, the group have found that 6 to 30% of 5' start sites of mouse and human transcripts initiate within repetitive elements. Remarkably, over 250,000 retrotransposon-derived transcription start sites show tissue- and cell-compartment-specific expression.

Leonard Lipovich (Wayne State University) and colleagues added 6,000 lncRNAs to this catalog by examining unclassified human cDNA clones and their expression profiles to determine whether these lncRNAs contribute to neurological disease phenotypes. They found that certain primate-specific and non-conserved lncRNAs are differentially expressed in brain regions that show high levels of activity. Some of these lncRNAs, antisense to protein-coding genes, can regulate their neighbors' expression. Weaving the intricacy of the transcriptome with the complexity of the mammalian body development and cognition, John Mattick (Garvan Institute of Medical Research) presented examples that emphasized the need to further understand the diversity of lncRNAs. Digging into the depths of the 'dark matter in the genome' using capture enrichment methods revealed not only numerous novel lncRNAs and their

*Correspondence: john_rinn@harvard.edu
[3]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138, USA
Full list of author information is available at the end of the article

isoforms but also isoforms of well-studied protein-coding mRNAs such as p53. Hundreds of lncRNAs were shown to change during stem cell differentiation and to have similar transcript stability to mRNAs, and many are associated with epigenetic complexes, suggesting that this complexity cannot be dismissed *en masse* as transcriptional noise.

RNA-RNA interactions

An ever emerging theme is the importance of RNA-RNA interactions and gene regulation. Kevin Morris (The Scripps Research Institute) described new findings on lncRNA-directed epigenetic regulation through RNA-RNA interactions. Morris and colleagues observed an antisense transcript from the *PTEN* pseudogene (*PTENpg1* asRNA), which is transcribed in the opposite direction to the previously reported *PTENpg1* sense transcript (which can sequester microRNAs and affects *PTEN* translation rates). The *PTENpg1* asRNA seems to direct transcriptional gene silencing of *PTEN* by interacting with the DNA methyltransferase Dnmt3a and the histone-lysine N-methyltransferase Ezh2 and affecting their localization to the *PTEN* promoter. Moreover, the *PTENpg1* asRNA (containing a poly(A) tail) seems to facilitate the cellular localization of *PTENpg1* sense transcript (lacking a poly(A) tail). Thus, this pseudogene node seems to control *PTEN* at both the translational and the transcriptional level.

ncRNAs in cellular and developmental biology

David Spector (Cold Spring Harbor Laboratory) and Shinichi Nakagawa (RIKEN Advanced Science Institute) presented a comprehensive investigation of the physiological roles of lncRNAs in mouse model systems. Using knockout mice, Spector and colleagues concluded that the lncRNA Malat1 was physiologically dispensable under laboratory conditions. This lack of phenotype may be attributed to genetic redundancy or stress conditions that were not examined. Indeed, Nakagawa examined the physiological role of the nuclear lncRNA *Neat1* (which is essential for the formation of nuclear structures involved in splicing termed paraspeckles) using knockout mice and observed a phenotype only in adult homozygote females, which became infertile at an earlier age. Seth Blackshaw (John Hopkins University School of Medicine) presented the *Six3OS* RNA as another example of a lncRNA required in developmental processes. Loss and gain of function of *Six3OS*, which is transcribed in the opposite direction from the Six3 transcription factor (TF) mRNA, showed that *Six3OS* regulates retinal cell specification. Several elegant follow-up experiments suggest that *Six3OS* regulates Six3-dependent transcription by recruiting the Polycomb repressive complex 2 (PRC2) to Six3 target genes.

Martin Walsh (Mount Sinai School of Medicine) presented his recent work on the complex interplay between RNA pol II and pol III, whereby RNA pol III can affect transcription by accentuating the function of RNA pol II. Ablation of the components of RNA pol III caused a significant depletion of an enhancer ncRNA in the *Sox2* locus followed by changes in chromatin signatures. Further analysis of the three-dimensional spacing of RNA pol III sites suggest that RNA pol III-regulated transcription could have a role in chromatin architecture, especially in the context of embryonic stem cell differentiation.

Rachel Duffié (Institut Curie) demonstrated how lncRNAs could control differential methylation patterns and imprinting during germ layer specification. Duffié presented a biallelically expressed lncRNA that was sufficient to induce DNA methylation *in cis*. Carla Klattenhoff (Massachusetts Institute of Technology) presented *braveheart*, a lncRNA required for stem cell differentiation into cardiomyocytes. Loss of *braveheart* causes disorganization and lack of beating in myofibrils and shorter sarcomeres. Collectively, these studies demonstrate the importance of lncRNAs in establishing cellular identity.

RNA and chromatin

The intersection of RNA with chromatin was introduced by the well-studied example of *Xist* by Neil Brockdorff (University of Oxford) and Jeannie Lee (Massachusetts General Hospital). Brockdorff showed that *Xist* RNA associates directly with additional proteins such as heterogeneous nuclear ribonucleoprotein U (hnRNPU) and PRC1. Brockdorff and colleagues have also identified a distinct form of PRC1, in which RING1 and YY1 binding protein (Rybp) replaces the catalytic Cbx subunit and associates with *Xist* RNA, suggesting that in addition to PRC2, hnRNPU and PRC1 are also involved in localizing *Xist* RNA to the inactive X. Lee further elucidated the mechanism by which the *Xist* lncRNA controls the silencing on the inactive X. Allele-specific localization profiles of PRC2 during X inactivation suggest a highly selective localization: first, PRC2 localizes to limited number of nucleation spots and, as *Xist* RNAs spreads *in cis*, PRC2 is recruited to additional sites in a gradient, unlike recruitment on autosomes.

As an attempt to generalize the observation that ncRNAs such as *Xist* bind the PRC2 and recruit it to its target loci, Richard Jenner (University College London) and colleagues observed short RNAs that are transcribed from CpG islands surrounding PRC2-bound promoters. Jenner's group showed that these short RNAs are required for efficient methylation of histone H3 on lysine 27 (H3K27me3) through interaction with a newly characterized amino-terminal domain of Suppressor of zeste 12

(Suz12). Collectively, these studies clearly demonstrate the interplay between DNA, lncRNAs, and proteins to establish proper epigenetic states and cellular physiology.

With the goal of identifying where RNA localizes on chromatin, Matthew Simon (Massachusetts General Hospital) and Jason West (Massachusetts General Hospital) introduced a new method (CHART) to purify regions directly bound by lncRNAs. CHART successfully identified the protein and chromatin targets of the paraspeckle-associated lncRNA *NEAT1* in human cells. Initial analysis of CHART coupled with sequencing revealed the chromosomal binding sites of *NEAT1* and suggested that *NEAT1* facilitates tertiary DNA interactions involved in nuclear architecture.

Deepening our understanding of recruitment of chromatin modifiers by lncRNAs, Howard Chang (Stanford University) and colleagues discovered the lncRNA *HOTTIP*, which activates the expression of its neighboring genes *in cis* by recruiting the activating modifier WDR5. Using crystallography and site-specific mutagenesis, the group identified the RNA-binding domain of WDR5. Moreover, binding to *HOTTIP* regulates the stability of WDR5, suggesting a model in which a RNA-based molecular clock regulates protein complex formation and stability. Similar to *HOTTIP*, Ramin Shiekhattar (Wistar Institute) presented a set of lncRNAs with enhancer-like functions. Shiekhattar described a specific example, in which an lncRNA enhances the expression of genes *in cis* through physical interactions with the Mediator complex. The lncRNA-Mediator interaction facilitates DNA looping to activate distal genes. Together, these suggest that lncRNAs emanating from enhancer regions may have RNA-based functions.

Using a detailed map of large intergenic ncRNAs (lincRNAs), one of us (JR) and colleagues used a computational 'guilt-by-association approach' to identify lincRNAs involved in specific pathways. One such example led to the discovery of a lncRNA, Required for adipogenesis 1 (*RAP1*). *RAP1* contains a striking feature: multiple repetitive exonic sequences that physically associate with hnRNPU. Single-molecule RNA localization methods showed that *RAP1* shows nuclear localization indicative of large chromosomal domains. Using a similar guilt-by-association approach, Keith Vance (University of Oxford) from Chris Ponting's group used computational and evolutionary genomics to identify mouse lincRNAs close to genes encoding developmental TFs. One such lincRNA is *Plinc*, which can act *in cis* to repress its bidirectionally promoted neighbor Pax6, as well as *in trans* to repress other targets involved in cell cycle and synaptic function. *Plinc* is enriched on chromatin and associates with the transcriptional cofactors Kap1 and Erh.

Non-mammalian ncRNA

ncRNAs also have critical regulatory roles in prokaryotes. Gisela Storz (National Institute of Child Health and Human Development) gave several examples from bacteria, in which small RNAs act in parallel to TFs to fine-tune the TFs' target gene levels through a base-pairing mechanism. In parallel, Jennifer Doudna (University of California, Berkeley) gave new insights into the CRISPR adaptive immune system in bacteria, in which viral DNA segments are integrated within bacterial DNA repeat clusters and are later transcribed to short RNAs that act in a viral defense pathway. Doudna highlighted some similarities between eukaryotic RNAi and prokaryotic CRISPR target recognition mechanisms based on molecular structures of these machineries.

RNA and disease

Andrew Feinberg (John Hopkins University School of Medicine) presented a comprehensive analysis of the epigenetic changes in cancer. Surveying differential DNA methylation in cancer revealed the loss of DNA methylation of the lncRNA termed *HOTS* (H19 opposite tumor suppressor) encoded antisense to the maternally imprinted *H19* gene, which encodes an untranslated RNA of unknown function. Feinberg described large lamin-associated DNA domains that are also lost in cancer. Similar to *HOTS*, almost all genes have naturally occurring antisense transcripts (NATs) that are ripe for RNA-RNA interactions and are often misregulated in disease states. Claes Wahlestedt (University of Miami) and colleagues have developed single-stranded oligonucleotides termed antagonists of NATs (antagoNATs) that upregulate specific endogenous genes *in vitro* and *in vivo* by interfering with the function of NATs. As NATs are widely expressed, the approach holds promise for potential RNA-based therapeutics. Together these studies highlight the need to understand the link between epigenetic regulation and lncRNAs in the etiology of disease.

Trans-generational RNA for the future

Paramutation is a form of non-Mendelian heredity in which the phenotypic expression of an allele is mutated while its genotype is intact. The meeting concluded with exciting updates on the current understanding of this phenomenon from worms to mammals. Andrew Spence (University of Toronto) described how a maternal transcript of the sex-determining gene *fem-1* does not require any coding capacity to trigger *fem-1* expression, which it does by preventing RNA-dependent epigenetic silencing of this gene. Oded Rechavi (Columbia University Medical Center) demonstrated how virus-derived small interfering RNAs can be trans-generationally transmitted in a non-Mendelian manner. Finally, Minoo Rassoulzadegan

(University of Nice) showed evidence for RNA-induced hereditary transmission of variation affecting the phenotype of three different genes in mouse animals. By electroporating an inducer small RNA homologous to the gene of interest to embryonic stem cells, Rassoulzadegan and colleagues were able to obtain a paramutable effect over many passages in cell culture. Collectively, these studies have provided exciting new evidence for the potential role of RNA in trans-generational inheritance in mammals.

This meeting was a resounding success and clearly demonstrated the versatility, diversity, and importance of ncRNAs in gene regulatory programs. John Mattick ended the meeting with inspiring words, highlighting the imminence of many groundbreaking discoveries to come from the dark matter of the genome. Kevin Morris reinforced the need for a collaborative spirit among the ncRNA community, which will rapidly drive the field to a greater success. One thing was clear: the transcriptome has revealed many mysterious creatures that comprise the mysterious world of 'Noncodarnia'.

Author details
[1]Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA. [2]Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, MA 02142, USA. [3]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138, USA. [4]Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA.

Published: 25 May 2012

doi:10.1186/gb-2012-13-5-315
Cite this article as: Hacisuleyman E, *et al.*: A Keystone for ncRNA. *Genome Biology* 2012, **13**:315.